

# Stochastic Ordering under Conditional Modelling of Extreme Values: Drug-Induced Liver Injury

Ioannis Papastathopoulos<sup>\*</sup> and Jonathan A. Tawn<sup>†</sup>

<sup>\*</sup>School of Mathematics, University of Bristol

<sup>†</sup>Department of Mathematics and Statistics, Lancaster University

i.papastathopoulos@bristol.ac.uk    j.tawn@lancaster.ac.uk

## Abstract

Drug-induced liver injury (DILI) is a major public health issue and of serious concern for the pharmaceutical industry. Early detection of signs of a drug's potential for DILI is vital for pharmaceutical companies' evaluation of new drugs. A combination of extreme values of liver specific variables indicate potential DILI (Hy's Law). We estimate the probability of severe DILI using the Heffernan and Tawn (2004) conditional dependence model which arises naturally in applications where a multidimensional random variable is extreme in at least one component. We extend the current model by including the assumption of stochastically ordered survival curves for different doses in a Phase 3 study.

**Keywords:** stochastic ordering; multivariate extremes; conditional dependence; drug toxicity; liver injury

## 1 Introduction

Drug-induced liver injury (DILI) is a major public health and industrial issue that has concerned clinicians for the past 50 years. The FDA (2008) reports that many drugs for a diverse range of diseases were either removed from the market or rejected at the pre-marketing stage because of severe DILI (e.g., iproniazid, ticrynafen, benoxaprofen, bromfenac, troglitazone, nefazodone, etc.). Therefore, signals of a drug's potential for DILI and early detection can help to improve the evaluation of drugs and aid pharmaceutical companies in their decision making. However,

in most clinical trials of hepatotoxic drugs, evidence of hepatotoxicity is very rare and although the pattern of injury can vary, there are no pathogenomic findings that make diagnosis of DILI certain even upon liver biopsy. Therefore, the majority of drugs that have been withdrawn fall into the post-marketing category and it is believed that the frequency of severe DILI does not exceed 0.01%.

Although the mechanism that causes DILI is not fully understood yet, the procedure at which its clinical assessment is performed stems from Zimmerman’s observation that hepatocellular injury sufficient to impair bilirubin excretion is a revealing indicator of DILI (Zimmerman 1978, 1999), also informally known as Hy’s Law. In other words, a finding of *alanine aminotransferase* (ALT) elevation, usually substantial, seen concurrently with *bilirubin* (TBL) greater than twice the upper limit of the normal (ULN), identifies a drug likely to cause severe DILI (fatal or requiring transplant) at a rate roughly 1/10 the rate of Hy’s Law cases. Moreover, these elevations should not be attributed to any other cause of injury; such as other drugs and *alkaline phosphatase* (ALP) should not be greatly elevated so as to explain TBL’s elevation.

Southworth and Heffernan (2012*b*) identified the assessment of DILI as an extreme value problem and using the Heffernan and Tawn (2004) modelling approach, analysed liver-related laboratory data. The use of the Heffernan and Tawn (2004) model in this context is supported by the flexibility of the model to allow a broad class of dependence structures and the possibility to describe the probabilistic behaviour of a vector random variable which is extreme in at least one margin. Despite its strong modelling potential, complications in terms of parameter identifiability problems and invalid inferences are experienced with the original modelling procedure of Heffernan and Tawn (2004). Keef et al. (2012) provided missing constraints for the parameter space of the Heffernan and Tawn (2004) model that are aimed to overcome these complications.

The data we consider in this study relates to observed liver-related variables from a sample of 606 patients who were issued a drug that has been linked to liver injury in a Phase 3 clinical trial and can be found in Southworth and Heffernan (2012*c*); see also Southworth and Heffernan (2012*b*). The patients were categorised into 4 different dose levels in a randomised, parallel group,

double blind Phase 3 clinical study. Our objective in this paper is to extend Keef et al. (2012) constraints to bound any two conditional distribution functions and hence incorporate scientific knowledge and efficiency into the modelling procedure of DILI by imposing the assumption of stochastically ordered survival curves in the joint tail region of the ALT and TBL variables between different dose levels of a drug. Therefore, estimation of ALT and TBL under the dose ordering assumption is beneficial as it sharpens inference and removes variability that arises in small sample sizes in the joint tail region of ALT and TBL that identifies DILI. Our motivation to impose additional constraints stems from the fact that the probability of DILI is logically ordered between different dose levels when the drug is liver toxic and this feature is unlikely to be evident from the data when considering clinical trials with small sample sizes as this particular one.

Table 1: Kendall's  $\tau$  estimates between ALT and TBL

	Baseline	Post-baseline	Residual
Dose A	0.13	0.25	0.12
Dose B	0.09	0.12	-0.01
Dose C	0.11	0.10	-0.02
Dose D	0.15	0.18	0.05

For example, the two central columns of Table 1 show the estimated Kendall's rank correlation of ALT and TBL measured at the baseline and post-baseline periods. The last column shows the Kendall's  $\tau$  of the residuals of ALT and TBL that were obtained from the median regression of the log-post-baseline on the log-baseline variable, see also Section 5.1. Table 1 conveys this lack of ordering since the estimated dependence of ALT and TBL at dose A appears to be significantly different from zero and even higher than any other dose in post-baseline and residual scale. On the other hand, the published literature reports jaundice, hepatitis and similar symptoms in approximately 1 out of 500 patients taking the dose D of this drug (Southworth and Heffernan, 2012b). We view such high dependencies in low doses as a by-product of sampling variability.

The proposed methodology and data analysis of the paper are based on an asymptotically motivated model of multivariate extreme value threshold model which is fitted to a fraction of the data. An alternative approach would be to model the joint distribution between the variables

using all the data through the use of empirically selected marginal and copula models (Joe, 1997; Nelsen, 2006). The former approach should exhibit less bias but larger variability than the latter, thus as the sample size increases the extreme value approach is likely to become more the efficient. The sample size in our study is probably about at the boundary where the extreme value methods have an advantage. Furthermore, even if the copula approach were to be adopted, the strategies developed here would still be relevant as the stochastic ordering issue would need addressing.

The paper is organised as follows. Section 2 describes the conditional dependence model of Heffernan and Tawn (2004) and the constraints of Keef et al. (2012). The additional constraints based on the assumption of stochastically ordered conditional distributions are presented in Section 3. The effect of the constraints of Keef et al. (2012) and this paper are assessed with a simulation study in Section 4. We illustrate in Section 5 the application of the additional constraints by analysing the multivariate extremes of ALT and TBL of the DILI data.

## 2 Methodology

### 2.1 Marginal transformation

Here and throughout vector algebra is applied componentwise. Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a continuous  $d$ -dimensional vector random variable and  $\Delta = \{1, \dots, d\}$ . We adopt the marginal transformation to approximate Laplace margins (Keef et al., 2012)

$$Y_i = \begin{cases} \log\{2\hat{F}_{X_i}(X_i)\} & \text{for } \hat{F}_{X_i}(X_i) < \frac{1}{2}, \\ -\log\{2[1 - \hat{F}_{X_i}(X_i)]\} & \text{for } \hat{F}_{X_i}(X_i) \geq \frac{1}{2}, \end{cases} \quad (1)$$

where the estimated distribution function  $\hat{F}_{X_i}$  is obtained from the semi-parametric model of Coles and Tawn (1994)

$$\hat{F}_{X_i}(x) = \begin{cases} 1 - \{1 - \tilde{F}_{X_i}(u_{X_i})\}\{1 + \hat{\xi}_i(x - u_{X_i})/\hat{\sigma}_i\}_+^{-1/\hat{\xi}_i} & \text{for } x > u_{X_i}, \\ \tilde{F}_{X_i}(x) & \text{for } x \leq u_{X_i}, \end{cases} \quad (2)$$

where  $\tilde{F}_{X_i}$  is the empirical distribution function and  $u_{X_i}$  is a threshold above which the gen-

eralised Pareto distribution with scale parameter and shape parameters  $\sigma_i$  and  $\xi_i$ , short-hand  $\text{GP}(\sigma_i, \xi_i)$ ,  $\sigma_i \in (0, \infty)$ ,  $\xi_i \in \mathbb{R}$ , is fitted to the observed values of the excess random variable  $X_i - u_{X_i} | X_i > u_{X_i}$  (Davison and Smith, 1990). The choice of the Laplace transformation is motivated by the symmetry of the Laplace distribution that ensures the limiting conditional dependence model to be unchanged for negatively dependent variables (Keef et al., 2012).

## 2.2 Conditional modelling of extreme values

The Heffernan and Tawn (2004) conditional dependence model characterises the probabilistic behaviour of the conditional vector random variable  $\mathbf{Y}_{-i} | Y_i = y$ , for large  $y$ , where  $\mathbf{Y}_{-i}$  denotes the  $(d-1)$ -dimensional vector of the transformed variables without the  $i$ -th margin. According to Heffernan and Tawn (2004), for each  $i \in \Delta$ , there exist vector-valued normalising functions,  $\mathbf{a}_{|i} : \mathbb{R} \rightarrow \mathbb{R}^{d-1}$  and  $\mathbf{b}_{|i} : \mathbb{R} \rightarrow \mathbb{R}^{d-1}$ , such that for  $x > 0$

$$\mathbb{P} \left\{ Y_i - u > x, \frac{\mathbf{Y}_{-i} - \mathbf{a}_{|i}(Y_i)}{\mathbf{b}_{|i}(Y_i)} \leq \mathbf{z} \middle| Y_i > u \right\} \rightarrow \exp(-x) G_{|i}(\mathbf{z}) \quad \text{as } u \rightarrow \infty, \quad (3)$$

where the  $j$ th marginal distribution  $G_{j|i}$  of  $G_{|i}$  is a non-degenerate distribution function for all  $j \in \Delta \setminus \{i\}$  and additionally, the following condition is required such that  $G_{|i}$  is well-defined:

$$\lim_{z \rightarrow \infty} G_{j|i}(z) = 1 \quad \text{for all } j \neq i$$

so there is no mass at  $+\infty$  but some is allowed at  $-\infty$ , in any margin. Heffernan and Tawn (2004) identified the normalising functions are unique up to type, and that for a broad class of distributions these functions are all in the parametric family

$$\mathbf{a}_{|i}(x) = \boldsymbol{\alpha}_{|i}x \quad \text{and} \quad \mathbf{b}_{|i}(x) = x^{\boldsymbol{\beta}_{|i}}, \quad x > 0,$$

with  $(\boldsymbol{\alpha}_{|i}, \boldsymbol{\beta}_{|i}) \in [-1, 1]^{d-1} \times (-\infty, 1)^{d-1}$ . Positive and negative dependence between variables  $Y_i, Y_j$ , for  $i \neq j$  is given by  $\alpha_{j|i} > 0$  and  $\alpha_{j|i} < 0$ , respectively, with  $\alpha_{j|i}$  the associated  $\boldsymbol{\alpha}_{|i}$  with  $Y_j$  variable. The strongest form of positive (negative) extremal dependence occurs when  $\alpha_{j|i} = 1$  ( $\alpha_{j|i} = -1$ ) and  $\beta_{j|i} = 0$  and is termed as asymptotic positive (negative) dependence, for all  $j \neq i$ . Otherwise, variables are termed asymptotically independent. The conditional model of Heffernan and Tawn (2004) can be viewed as a multivariate semiparametric regression of  $\mathbf{Y}_{-i}$

on  $Y_i$ , i.e. given  $Y_i > u$ , for large  $u$  we have:

$$\mathbf{Y}_{-i} = \boldsymbol{\alpha}_{|i}x + x^{\boldsymbol{\beta}_{|i}}\mathbf{Z}_{|i} \quad \text{for } Y_i = x > u \quad (4)$$

where  $\mathbf{Z}_{|i}$  is a  $d - 1$  dimensional variable with non-zero mean and distribution function  $G_{|i}$ . The original procedure of Heffernan and Tawn (2004) for estimating the vector parameters  $\boldsymbol{\alpha}_{|i}$  and  $\boldsymbol{\beta}_{|i}$  consists of using pseudo-likelihood methods to jointly estimate the parameters of interest. In particular, if we assume that  $\mathbf{Z}_{|i}$  has finite vector mean  $\boldsymbol{\mu}_{|i}$  and standard deviations  $\boldsymbol{\sigma}_{|i}$ , then the mean and standard deviation of the conditional random variable  $\mathbf{Y}_{-i}|Y_i > u$  is  $\boldsymbol{\alpha}_{|i}Y_i + \boldsymbol{\mu}_{|i}(Y_i)^{\boldsymbol{\beta}_{|i}}$  and  $Y_i^{\boldsymbol{\beta}_{|i}}\boldsymbol{\sigma}_{|i}$ , respectively. Under the false working assumption that  $\mathbf{Z}_{|i}$  are independent Normal random variables, numerical maximisation of the likelihood of the model over the parameter space is required to get parameter estimates  $(\hat{\boldsymbol{\alpha}}_{|i}, \hat{\boldsymbol{\beta}}_{|i}, \hat{\boldsymbol{\mu}}_{|i}, \hat{\boldsymbol{\sigma}}_{|i})$ , and  $G_{|i}$  is estimated nonparametrically by the empirical distribution function of:

$$\hat{\mathbf{Z}}_{|i} = \frac{\mathbf{Y}_{-i} - \hat{\boldsymbol{\alpha}}_{|i}Y_i}{(Y_i)^{\hat{\boldsymbol{\beta}}_{|i}}}. \quad (5)$$

Once estimates are obtained, standard procedures for inference and extrapolation can be performed as in Heffernan and Tawn (2004) by implementing Algorithm 1. As an example, the functional  $\mathbb{P}(\mathbf{X} \in C | X_i > s)$  can be approximated by repeating steps 1–5, and evaluating the estimate as the long run proportion of the generated sample that falls in a set  $C \in \mathbb{R}^d$ . As far as the confidence intervals of the estimate of any functional are concerned, these are obtained by the replication of the three stages of the following bootstrap method: data generation under the fitted model, estimation of model parameters and the derivation of an estimate of any derived parameters linked to extrapolation.

---

**Algorithm 1** Sampling Algorithm

---

- 1: Simulate  $Y_i$  from the Laplace distribution conditional on its exceeding threshold  $u > 0$ .
  - 2: Sample  $\mathbf{Z}_{|i}$  from  $\hat{G}_{|i}$  independently of  $Y_i$ .
  - 3: Obtain  $\mathbf{Y}_{-i} = \boldsymbol{\alpha}_{|i}Y_i + (Y_i)^{\boldsymbol{\beta}_{|i}}\mathbf{Z}_{|i}$ .
  - 4: Transform  $\mathbf{Y} = (\mathbf{Y}_{-i}, Y_i)$  to the original scale by using the inverse transformation of (1) for each margin.
  - 5: The resulting transformed vector  $\mathbf{X}$  constitutes a simulated value from the conditional distribution of  $\mathbf{X}|X_i > t^{-1}(u)$ , where  $t^{-1}(\cdot)$  denotes the inverse transformation of (1).
-

### 2.3 Inference based on Keef et al. (2012)

Although the efficiency of the model has led to its implementation in a wide range of applications including riverflow and rainfall (Keef et al., 2009), temporal river flow cases (Eastoe and Tawn, 2012), food safety (Paulo et al., 2006) and finance (Abbas et al., 2011), it was recently discovered by Keef et al. (2012) that further constraints on the parameter space of the model are required. According to Sibuya (1960) and Tiago de Oliveira (1962/63), there are different categorisations of extremal dependence between two random variables  $(X_i, X_j)$ , i.e. asymptotic dependence and asymptotic independence measured by the coefficients of tail dependence

$$\chi_{ij}^+ = \lim_{p \rightarrow 1} \mathbb{P} \left\{ X_j > F_j^{-1}(p) | X_i > F_i^{-1}(p) \right\}, \quad (6)$$

$$\chi_{ij}^- = \lim_{p \rightarrow 1} \mathbb{P} \left\{ X_j < F_j^{-1}(1-p) | X_i > F_i^{-1}(p) \right\}.$$

When  $\chi_{ij}^+ > 0$  ( $\chi_{ij}^- > 0$ ) the variables are termed asymptotically positive (negative) dependent and asymptotically independent, otherwise. Taking these measures into consideration, Heffernan and Tawn (2004) omitted the fact that there is stochastic ordering between asymptotically independent and dependent models. In particular, let the  $q$ th conditional quantile of  $Y_j | Y_i = x$ , for large  $x$  under the Heffernan and Tawn (2004) model be  $y_{j|i}(q) = \alpha_{j|i}x + x^{\beta_{j|i}}z_{j|i}(q)$ , the associated quantile under asymptotic positive dependence  $y_{j|i}^+(q) = x + z_{j|i}^+(q)$  and the associated quantile under asymptotic negative dependence  $y_{j|i}^-(q) = -x + z_{j|i}^-(q)$ . The natural restriction

$$y_{j|i}^-(q) \leq y_{j|i}(q) \leq y_{j|i}^+(q), \quad \forall q \in [0, 1], \quad (7)$$

imposes further constraints on the parameter space of the model which are given by (Theorem 1.1, Keef et al. (2012)):

Case I: either

$$\alpha_{j|i} \leq \min \left\{ 1, 1 - \beta_{j|i} z_{j|i}(q) v^{\beta_{j|i}-1}, 1 - v^{\beta_{j|i}-1} z_{j|i}(q) + v^{-1} z_{j|i}^+(q) \right\}.$$

or

$$1 - \beta_{j|i} z_{j|i}(q) v^{\beta_{j|i}-1} < \alpha_{j|i} \leq 1 \text{ and } (1 - \beta_{j|i}^{-1}) \{\beta_{j|i} z_{j|i}(q)\}^{1/(1-\beta_{j|i})} (1 - \alpha_{j|i})^{-\beta_{j|i}/(1-\beta_{j|i})} + z_{j|i}^+(q) > 0.$$

Case II: either

$$-\alpha_{j|i} \leq \min \left\{ 1, 1 + \beta_{j|i} v^{\beta_{j|i}-1} z_{j|i}(q), 1 + v^{\beta_{j|i}-1} z_{j|i}(q) - v^{-1} z_{j|i}^-(q) \right\}$$

or

$$1 + \beta_{j|i} v^{\beta_{j|i}-1} z_{j|i}(q) < -\alpha_{j|i} \leq 1 \text{ and } (1 - \beta_{j|i}^{-1}) (-\beta_{j|i} z_{j|i}(q))^{1/(1-\beta_{j|i})} (1 + \alpha_{j|i})^{-\beta_{j|i}/(1-\beta_{j|i})} - z_{j|i}^-(q) > 0.$$

where  $v > u$  is a value above the maximum observed value of  $Y_i$  so that the constraints are imposed only on extrapolations. As far as the selection of  $q$  is concerned, Keef et al. (2012) found that for both cases the conditions were satisfied for all  $q$  if they were each satisfied for both  $q = 0$  and  $q = 1$ .

### 3 Estimation of Heffernan and Tawn (2004) model under stochastic ordering

#### 3.1 Quantile Ordering Constraints

In this paper we exploit the same idea for the construction of the parameter space of Heffernan and Tawn (2004) model under the assumption of stochastic ordering between two conditional random variables. Specifically, let the  $q$ th,  $q \in [0, 1]$ , conditional quantile of  $Y_l|Y_j = x$  and  $Y_k|Y_i = x$  for large  $x$  be  $y_{l|j}(q)$  and  $y_{k|i}(q)$ , respectively. Under the Heffernan and Tawn (2004) model we have that  $y_{l|j}(q) = \alpha_{l|j}x + x^{\beta_{l|j}} z_{l|j}(q)$  and  $y_{k|i}(q) = \alpha_{k|i}x + x^{\beta_{k|i}} z_{k|i}(q)$ . Our objective is to derive the constraints under which there is stochastic ordering between the conditional variables so that the following condition is always satisfied for all  $x$  above a level  $v > u$

$$y_{k|i}(q) \leq y_{l|j}(q), \quad \forall q \in [0, 1]. \quad (8)$$



The motivation for exploring inequality (8) stems from the dose ordering effect in the joint region of ALT and TBL. Consider for example the transformed with respect to equation (1) ALT and TBL and let  $y_{2|1}^A(q)$  and  $y_{2|1}^B(q)$  be the conditional quantiles of TBL given a large level of ALT for dose  $A$  and  $B$ , respectively. Then under the assumption of liver toxicity, it is intuitive to consider the natural ordering of the conditional quantiles  $y_{2|1}^A(q) < y_{2|1}^B(q)$ . The following theorem gives conditions under which two conditional quantiles based on Heffernan and Tawn (2004) satisfy the ordering constraint (8), for a  $q \in [0, 1]$ .

**Theorem 1.** *Let  $D(x) : [v, \infty) \rightarrow \mathbb{R}$  such that  $D(x) := (\alpha_{l|j} - \alpha_{k|i})x + x^{\beta_{l|j}} z_{l|j}(q) - x^{\beta_{k|i}} z_{k|i}(q)$ , with  $(\alpha_{l|j}, \alpha_{k|i}) \in [-1, 1]^2$ ,  $(\beta_{l|j}, \beta_{k|i}) \in (-\infty, 1)^2$ , and  $(z_{l|j}, z_{k|i}) \in \mathbb{R}^2$ . For  $v \geq u$  and  $q \in [0, 1]$  i) the stationary points (s.p.) of  $D(x)$  can be categorised as in Table 2*

Table 2: Conditions for stationary points of  $D(x)$ .

number of s.p.	$D'(v)$	$s$	$D'(s), s \in \mathbb{R}$
0	$> 0$	complex/real	$(0, \infty)$
1	$< 0$	complex/real	$(-\infty, \infty)$
2	$> 0$	real	$(-\infty, 0)$

where

$$s = \left\{ \frac{\beta_{k|i}(\beta_{k|i} - 1)z_{k|i}(q)}{\beta_{l|j}(\beta_{l|j} - 1)z_{l|j}(q)} \right\}^{1/(\beta_{l|j} - \beta_{k|i})} \in \mathbb{C},$$

and  $\mathbb{C}$  denotes the set of complex numbers,

ii) The ordering constraint (8) holds for all  $x > v$  if  $\alpha_{l|j} \geq \alpha_{k|i}$  and either

1. when  $D(x)$  has no s.p.:  $D(v) \geq 0$ , or
2. when  $D(x)$  has one s.p.  $x_*$ :  $\min \{D(v), D(x_*)\} \geq 0$ , or
3. otherwise:  $\min \{D(v), D(x_*), D(x_{**})\} \geq 0$ , where  $x_*$  and  $x_{**}$  are the two s.p..

## Proof

(i) According to Descartes's rule of signs  $D'(x) = 0$  can have at most two solutions, therefore  $D(x)$  can have at most two stationary points. Numerical inspection of the function (e.g. for  $\alpha_{l|j} = 0.2, \alpha_{k|i} = 0.1, \beta_{l|j} = 0.2, \beta_{k|i} = 0.5, z_{l|j}(q) = 0.6$  and  $z_{k|i}(q) = 0.6$ ) shows that there can be cases where  $D(x)$  has two stationary points. The cases of Table 2 follow from noting that  $D''(s) = 0$  is the unique root of  $D''(x)$ , so that  $D'(x)$  has at most one stationary point, i.e.

when  $s > v \in \mathbb{R}$  then  $s$  is a s.p. of  $D'(x)$ , otherwise  $s$  is a complex number so that  $D'(x)$  is monotone for  $x > v$ .

(ii)  $D(x) \geq 0$  for  $x \in [v, \infty)$ , implies that  $\lim_{x \rightarrow \infty} D(x)$  can be either 0 or  $\infty$ , so that:

$$\alpha_{l|j} \geq \alpha_{k|i}. \quad (9)$$

Categorising the cases with respect to the number of stationary points of  $D(x)$ , we have that  $D(x) \geq 0$ , for all  $x > v$ , if and only if one of the 3 conditions of Theorem 1 (ii) holds.  $\square$

From a statistical perspective, constraints follow from the nature of the  $D(x)$  function, i.e. one needs to find the stationary points of  $D(x)$  so that estimation of Heffernan and Tawn (2004) parameters under the quantile ordering assumption can be carried out. Since the function  $D'(x)$  is not linear, closed form roots of  $D'(x) = 0$  do not exist. Theorem 1 (i) gives the number of s.p. of  $D(x)$  together with the intervals that these points can lie, e.g. if  $D(x)$  has one s.p. then one dimensional root finding is sufficient to estimate the root of  $D'(x)$  and if  $D(x)$  has two s.p. the domain of the function  $D(x)$  can be separated into two subintervals  $(v, s)$  and  $(s, \infty)$  both in which one dimensional root finding is sufficient to yield estimates of these two s.p..

### 3.2 Inference based on stochastic ordering assumption

As far as the estimation of the Heffernan and Tawn (2004) model under stochastic ordering is concerned, Theorem 1 provides a set of exclusive cases where each one shows the number of stationary points that the function  $D(x)$  can have. This provides an automatic way for selecting the associated stochastic ordering constraint from Theorem 1 under which maximisation of the likelihood of the model is performed. For the required stochastic ordering (8) constraint we found that the conditions of Theorem 1 were satisfied for all  $q$  if they were satisfied for both  $q = 0$  and 1. To illustrate the feature Figure 1 shows the profile loglikelihood surface for combinations of the parameters of the conditional dependence model parameters of TBL given ALT for dose  $A$ , denoted by  $\alpha_{2|1}^A$  and  $\beta_{2|1}^A$ , under the assumption that the conditional quantile of dose  $A$  is smaller than the conditional quantile of dose  $B$ . Solid lines correspond to the joint  $q = 0$  and  $q = 1$  constraints whereas dashed lines correspond to  $0 < q < 1$  constraints. The parameter space obtained from the joint  $q = 0$  and  $q = 1$  constraints is nested in the parameter spaces

obtained from  $0 < q < 1$  constraints.

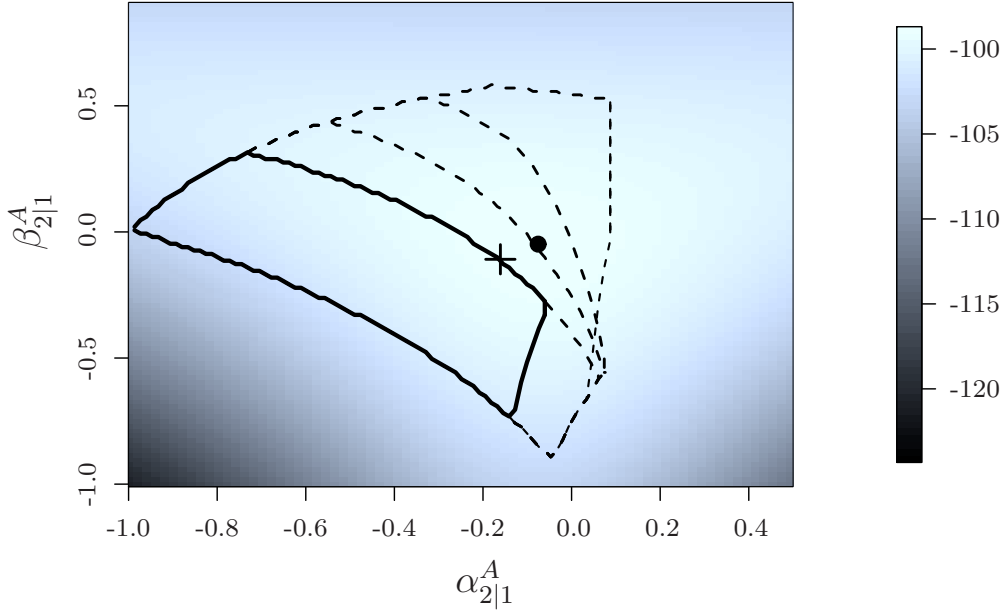


Figure 1: Profile log-likelihood surface for dose  $A$  parameters  $(\alpha_{2|1}^A, \beta_{2|1}^A)$ . The solid curves show the boundary of the parameter space under the constraints of Theorem 1 when  $q = 0$  and  $q = 1$ . Dashed curves show the constraints of Theorem 1 when  $0 < q < 1$ , showing these constraints are less restrictive than when  $q = 0$  and  $q = 1$ . The dot and cross show estimated parameters for unconstrained and constrained estimation respectively.

## 4 Simulation study of ordering constraints

The impact of the proposed constraints of Section 3.1 and the Keef et al. (2012) constraints is illustrated with a simulation study. We examine the performance of conditional quantile estimates using simulated datasets from three bivariate copula models with Laplace marginals. Denote by  $(Y_1, Y_2)$  a random variable arising from one of these copula models. We simulate pairs of observations conditionally on  $Y_1$  exceeding a finite threshold  $u$  from the *exact* form of the limiting conditional dependence model. Explicitly, we assume that the conditional distribution function of  $Y_2|Y_1 > u$ , for finite  $u$ , is equal to the actual limiting distribution function that is implied by expression (3) for each of the three copula models of Section 4.1, i.e., we assume that

$$Y_2 = \alpha_{2|1} Y_1 + Y_1^{\beta_{2|1}} Z_{2|1}, \quad Z_{2|1} \sim G_{2|1}, \quad Y_1 > u \quad \text{and} \quad Z_{2|1} \text{ independent of } Y_1 \quad (10)$$

with  $\alpha_{2|1}$ ,  $\beta_{2|1}$  and  $G_{2|1}$  chosen such that expression (3) holds. In Section 4.1 we present the three copula models along with their Heffernan and Tawn (2004) limiting representation and the simulation design. Algorithm 2 provides the simulation procedure used in this study for the bivariate case. In Section 4.2 we discuss the results of the simulation study.

---

**Algorithm 2** Simulation

---

- 1: Set  $I = 1$ ,  $(\alpha_{2|1}, \beta_{2|1}) \in (-1, 1) \times (-\infty, 1)$  and  $N \in \mathbb{N}$ .
  - 2: Simulate  $y_{1,I}$  from a Laplace distribution conditional on exceeding a threshold  $u$ .
  - 3: Simulate  $z_{2|1,I}$  independently of  $y_{1,I}$  from the true limiting distribution  $G_{2|1}$
  - 4: Set  $y_{2,I} = \alpha_{2|1}y_{1,I} + y_{1,I}^{\beta_{2|1}} z_{2|1,I}$ .
  - 5: If  $I < N$  set  $I = I + 1$  and go to step 2; otherwise return  $(\mathbf{y}_1, \mathbf{y}_2)$ , where  $\mathbf{y}_1 = (y_{1,1}, \dots, y_{1,N})'$ ,  $\mathbf{y}_2 = (y_{2,1}, \dots, y_{2,N})'$ .
- 

## 4.1 Theoretical models

The three copula models we consider in the simulation study are the bivariate extreme value logistic, the inverted logistic and the Gaussian with distribution functions given respectively by

$$\mathbb{P}(Y_1 \leq x, Y_2 \leq y) = \exp \left( - \left[ \{-\log F(x)\}^{1/\kappa} + \{-\log F(y)\}^{1/\kappa} \right]^\kappa \right), \quad \kappa \in (0, 1],$$

$$\mathbb{P}(Y_1 \leq x, Y_2 \leq y) = \mathbb{P}(X > -x, Y > -y), \quad \text{where } (X, Y) \sim \text{logistic}(\kappa),$$

$$\mathbb{P}(Y_1 \leq x, Y_2 \leq y) = \Phi_\Sigma \left[ \Phi^{-1} \{F(x)\}, \Phi^{-1} \{F(y)\} \right],$$

where  $F$ ,  $\Phi$  and  $\Phi_\Sigma$  denote the d.f. of the standard Laplace, the standard normal and the standard bivariate Gaussian with correlation matrix  $\Sigma = (\sigma_{i,j})$  with  $\sigma_{i,i} = 1$  and  $\sigma_{i,j} = \rho$ ,  $\rho \in (-1, 1)$ , for  $i \neq j$ . The normalising functions and the residual distribution  $G_{2|1}$  of the limiting representation (3) are summarised in Table (3)

The values of the parameters  $\kappa$  and  $\rho$  used in the simulation study are chosen such that the simulated data preserve the stochastic ordering feature. For example, in the logistic copula case, dependence increases as the value of  $\kappa$  decreases which implies larger joint survival probabilities as  $\kappa$  decreases. We thus simulate pairs of observations from the exact conditional dependence model (10) under  $\kappa = 0.6$  and  $\kappa = 0.9$ . For the asymptotically independent models, two pairs

Table 3: Heffernan and Tawn (2004) normalising constants  $\alpha_{2|1}$ ,  $\beta_{2|1}$ , and limiting distribution  $G_{2|1}$  for the bivariate extreme value logistic, inverted logistic and standard bivariate Gaussian copula with correlation parameter  $\rho > 0$ .

Model	$\alpha_{2 1}$	$\beta_{2 1}$	$G_{2 1}(z)$
logistic	1	0	$(1 + \exp\{-z/\kappa\})^{\kappa-1}$
inverted logistic	0	$1 - \kappa$	$1 - \exp(-\kappa z^{1/\kappa})$
Gaussian	$\rho^2$	$1/2$	$N(0, 2\rho^2(1 - \rho^2))$

of parameter values are used in the simulation study, i.e., for the inverted logistic copula we use  $\kappa = (0.3, 0.7)$  and  $\kappa = (1, 0.415)$ , and for the Gaussian copula we use  $\rho = (0.3, 0.7)$  and  $\rho = (0, 0.5)$ . The second pair of  $\kappa$  and  $\rho$  values is used for comparisons between the two asymptotically independent models as a key summary measure of extremal tail dependence, the coefficient of tail dependence of Ledford and Tawn (1996), between the variables  $Y_1$  and  $Y_2$  is the same for these choices.

The conditional quantile estimates are obtained from the original Heffernan and Tawn (2004) model, the constrained model of Keef et al. (2012) and the constrained model described in Section 3.1. We refer to these models as HT, AD and SO, respectively. On this note, the stochastic ordering constraints are imposed on pairs of observations with different dependence parameters. The performance of the estimates is assessed with the Monte Carlo estimate of the root mean square error. To be specific let  $y(q)$  and  $\hat{y}(q)$  be the true conditional quantile and its model-based estimate. The Monte Carlo estimate of the root mean square error is

$$\widehat{\text{RMSE}}(q) = \sqrt{\frac{1}{m} \sum_{i=1}^m \{\hat{y}_i(q) - y(q)\}^2},$$

where  $\hat{y}_1(q), \dots, \hat{y}_m(q)$  denotes a Monte Carlo sample of the conditional quantile estimates and each estimate is obtained from a simulated sample of  $N \in \mathbb{N}$  pairs of observations,  $\{(\mathbf{y}_1, \mathbf{y}_2) : y_{1,i} > u, i = 1, \dots, N\}$ . For all three models, the root mean square error is estimated over a grid values of  $q \in [0, 1]$  and three conditioning levels, i.e.,  $x_{0.95}$ ,  $x_{0.99}$  and  $x_{0.999}$ , where  $x_p$  denotes the  $p$ -th quantile of the standard Laplace distribution. Comparisons are made on the basis of ratios of RMSEs between estimates from different models. For the constrained models we tabulate the percentage of the Monte Carlo samples where estimates changed with respect

to the original Heffernan and Tawn (2004) model.

## 4.2 Results of simulation study

Table 4 shows the percentage of the estimates that changed with respect to one of the three reference models (HT, AD, SO). The imposition of constraints to the parameter space of the HT model alters estimates particularly in the asymptotically independent models and less in the asymptotically dependent model. In particular, the larger changes occur when variables are highly dependent except for the logistic model. Regarding the logistic model, the percentage of changes in the first two rows is relatively small compared to the other models. This feature is explained by the greater concentration of the maximum likelihood estimates around the true value, which lies in the feasible set of the constrained space (Keef et al., 2012). Additionally, small changes occur within the asymptotically independent models especially when the variables do not possess strong dependence. For the second pair of parameter values for the inverted logistic and Gaussian copulas, the changes in parameter estimates do not occur at a similar rate when dependence between variables is present ( $\kappa = 0.415$ ,  $\rho = 0.5$ ). We therefore conclude that the constraints induced by the AD and SO models are not only related to the level of dependence but to the dependence structure as well. Table 5 shows the ratio of the Monte Carlo root mean

Table 4: Percentage of estimates that changed with respect to a reference model. First row: percentage of AD estimates different from the HT estimates, second row: percentage of SO estimates different from the HT estimates, third row: percentage of SO estimates different from AD estimates. Columns show the corresponding copula model from Section 4.1 used in the simulation.

	logistic copula		inverted logistic copula				Gaussian copula			
$\kappa$ or $\rho$	0.6	0.9	0.3	0.7	0.415	1	0.7	0.3	0.5	0
AD-HT	29%	27%	63%	10%	41%	0.3%	36%	0%	6%	1%
SO-HT	54%	56%	77%	45%	68%	47%	42%	10%	30%	25%
SO-AD	33%	46%	43%	44%	47%	47%	9%	9%	24%	25%

square error, of the conditional quantile estimates obtained from the three copula models. An increase in efficiency under the imposition of the constrained models AD and SO is observed for nearly all conditional quantile estimates in the asymptotically independent models. The highest reduction in RMSE is achieved by the SO model in the inverted logistic copula, a feature which is also consistent with the higher percentage of change in estimates as shown

in Table 4. The conclusion for the asymptotically independent models is that the efficiency of the conditional quantile estimates is, in decreasing order, SO, AD and HT. Regarding the asymptotically dependent logistic copula, constrained models appear to be less efficient than the HT model and the efficiency of the conditional quantile estimates is, in decreasing order, HT, AD and SO.

Table 5: Ratio of the Monte Carlo root mean square error of the conditional quantile estimates  $\hat{y}_{2|1}(q)$  obtained from the HT, AD and SO models. Results are reported for  $q = 0.2, 0.5, 0.8$  and three conditioning levels  $x_{0.95}$ ,  $x_{0.99}$  and  $x_{0.999}$ . The value  $x_p$  here denotes the  $p$ -th quantile of the standard Laplace distribution.

	logistic copula			inverted logistic copula				Gaussian copula			
$\kappa$ or $\rho$	0.6	0.9		0.3	0.7	0.415	1	0.7	0.3	0.5	0
$q = 0.2$											
AD/HT	$x_{0.95}$	1.10	1.02	0.99	1.00	0.98	1.00	1.00	1.00	0.99	0.99
	$x_{0.99}$	1.00	1.08	0.99	0.97	0.98	0.98	1.00	1.00	1.00	1.00
	$x_{0.999}$	1.10	1.04	1.00	0.96	0.97	0.98	0.99	1.00	0.99	0.99
SO/HT	$x_{0.95}$	1.20	0.97	0.93	0.99	0.97	0.96	0.98	0.99	0.99	1.00
	$x_{0.99}$	0.99	1.24	0.95	0.97	0.94	0.96	0.97	0.99	0.99	1.00
	$x_{0.999}$	0.96	1.12	0.96	0.94	0.94	0.98	0.97	0.99	0.97	0.97
SO/AD	$x_{0.95}$	1.10	0.94	0.93	0.99	0.99	0.96	0.98	0.99	0.99	1.00
	$x_{0.99}$	0.96	1.14	0.95	0.99	0.95	0.97	0.97	0.99	0.99	1.00
	$x_{0.999}$	1.01	1.07	0.96	0.98	0.96	0.99	0.98	0.99	0.97	0.97
$q = 0.5$											
AD/HT	$x_{0.95}$	0.99	1.05	1.02	1.02	0.98	1.00	1.00	1.00	0.99	1.00
	$x_{0.99}$	1.09	1.08	1.01	0.96	0.98	0.98	1.01	1.00	1.00	1.00
	$x_{0.999}$	1.11	1.08	1.00	0.89	0.94	0.95	1.00	1.00	0.99	0.97
SO/HT	$x_{0.95}$	0.78	1.43	0.85	1.02	0.89	0.98	0.94	0.99	0.94	1.01
	$x_{0.99}$	1.09	1.29	0.79	0.96	0.82	0.93	0.90	0.98	0.93	1.00
	$x_{0.999}$	1.15	1.27	0.84	0.85	0.82	0.92	0.86	0.98	0.90	0.90
SO/AD	$x_{0.95}$	0.78	1.35	0.84	0.99	0.89	0.97	0.94	0.99	0.94	1.01
	$x_{0.99}$	0.99	1.19	0.78	0.99	0.83	0.94	0.89	0.98	0.93	1.00
	$x_{0.999}$	1.03	1.17	0.84	0.95	0.86	0.96	0.86	0.98	0.91	0.92
$q = 0.8$											
AD/HT	$x_{0.95}$	1.04	1.05	1.02	1.03	1.00	0.99	1.00	1.00	0.99	1.00
	$x_{0.99}$	1.10	1.08	1.03	0.98	0.99	0.99	1.01	1.00	1.00	0.99
	$x_{0.999}$	1.14	1.09	0.81	0.86	0.94	0.93	1.00	1.00	0.99	0.95
SO/HT	$x_{0.95}$	0.97	1.25	0.84	1.04	0.84	0.99	0.91	0.99	0.92	1.02
	$x_{0.99}$	1.12	1.29	0.73	0.99	0.77	0.95	0.87	0.98	0.90	0.99
	$x_{0.999}$	1.18	1.31	0.71	0.82	0.75	0.88	0.81	0.97	0.88	0.88
SO/AD	$x_{0.95}$	0.93	1.19	0.81	1.00	0.84	1.00	0.91	0.99	0.92	1.01
	$x_{0.99}$	1.01	1.19	0.71	1.00	0.78	0.95	0.86	0.98	0.90	1.00
	$x_{0.999}$	1.04	1.19	0.74	0.94	0.79	0.94	0.80	0.97	0.88	0.92

## 5 Application: drug-induced liver injury

### 5.1 Preprocessing and outline of analysis

The data that we consider in this study relates to a sample of 606 patients that were issued a drug linked to liver injury in a randomised, parallel group, double blind Phase 3 clinical study. ALT and TBL measurements were collected from all patients at baseline (prior to treatment) and post-baseline (after 6 weeks of treatment) periods. Let  $V_{i,B}^j$  and  $V_{i,P}^j$  be the  $i$ -th baseline and post-baseline laboratory variable respectively, measured at dose  $j = A, B, C$  and  $D$ . We use  $i = 1, 2$  to denote the ALT and TBL, respectively.

Instead of working with the raw data, the Box and Cox (1964) transformation is applied initially to stabilise the heterogeneity observed in the samples. For this dataset we apply the log-transformation and we denote the transformed data by  $W_{i,B}^j = \log(V_{i,B}^j)$  and  $W_{i,P}^j = \log(V_{i,P}^j)$ . Consequently, we use a robust linear regression model of the log-post-baseline on the log-baseline variable to adjust for the baseline effect, i.e., in its simplest form, the robust linear regression of  $W_{i,P}^j$  on  $W_{i,B}^j$  admits

$$W_{i,P}^j = \gamma_i^j + \delta_i^j W_{i,B}^j + X_i^j, \quad i = 1, 2, \quad j = A, \dots, D, \quad (11)$$

where  $(\gamma_i^j, \delta_i^j) \in \mathbb{R}^2$  and  $X_i^j$  is a zero mean error random variable. Here we use median quantile regression (Koenker and Bassett, 1978) which is equivalent to assuming that the error random variable  $X_i^j$  follows the Laplace distribution with zero location constant scale parameters (Yu and Moyeed, 2001). Our approach is based on the basic model structure of Southworth and Heffernan (2012b), i.e., the extremal dependence of  $(X_1^j, X_2^j)$  is estimated from the Heffernan and Tawn (2004) conditional dependence model whereas the log-baseline variables  $W_{1,B}^j$  and  $W_{2,B}^j$  are modelled independently for each dose. Under the assumption of independence between  $X_i^j$  and  $W_{i,B}^j$  simulated samples of the post-baseline variables can be generated. The exact procedure of the simulation is straightforward, i.e., residual and baseline samples are generated from their models and are combined in equation (11), with  $\gamma_i^j$  and  $\delta_i^j$  replaced by their corresponding maximum likelihood estimates, to produce simulated samples for the log-post-baseline variable  $W_{i,P}^j$ . The simulated sample is then back-transformed to its original scale using the inverse



Box-Cox transformation.

The key differences between our modelling procedure and Southworth and Heffernan (2012*b*) are related to the modelling of the baseline and the estimation of the conditional dependence model parameters. Firstly, for each baseline variable we implement the univariate semi-parametric model of Coles and Tawn (1994) as described in Section 2.1 by equation (2) whereas Southworth and Heffernan (2012*b*) use the empirical distribution function. Our motivation for modelling the tail of the baseline variable stems from the fact that it is likely to observe higher baseline ALT and TBL in the population (post-marketing period) than in the clinical trial (pre-marketing period). Therefore, tail modelling of the baseline is key to the simulation process as it incorporates a natural source of extremity through model-based extrapolation. Results from the univariate analysis are not presented in this paper but similar analyses can be found in Southworth and Heffernan (2012*a*) and Papastathopoulos and Tawn (2012). As far as the estimation of the conditional dependence model's parameters is concerned, we impose the stochastic ordering constraints developed in Section 3 and compare the results with the unconstrained estimates obtained from the HT model.

In Section 5.2 we illustrate the effect of the stochastic ordering constraints via estimates of conditional quantiles for all doses and in Section 5.3 we proceed to the prediction of the probability of DILI by simulating post-baseline laboratory data of hypothetical populations of size 200000 using the fitted marginal and conditional dependence models. The assessment of the uncertainty of the estimates of extreme quantities is performed via the bootstrap procedure.

## 5.2 Dependence modelling

Let  $Y_1^j$  and  $Y_2^j$  be the transformed, with respect to equation (1), residuals  $X_1^j$  and  $X_2^j$  for each dose  $j$ . Figure 2 shows the bivariate scatterplots of  $Y_2^j$  against  $Y_1^j$  for all dose levels. The dependence between the residual ALT and TBL variables appears to be very weak for all dose levels and a direct conclusion regarding the stochastic ordering effect cannot be made on the basis of Figure 2. A lack of ordering appears from the estimated conditional quantiles of TBL given ALT from the HT model as shown in Figure 3 in the standard Laplace scale. The estimates

of the median conditional quantiles from the HT model are ordered above approximately the conditioning level 4 whereas the minimum and maximum conditional quantile estimates exhibit a lack of ordering for the majority of the conditioning levels. The effect of constraining the

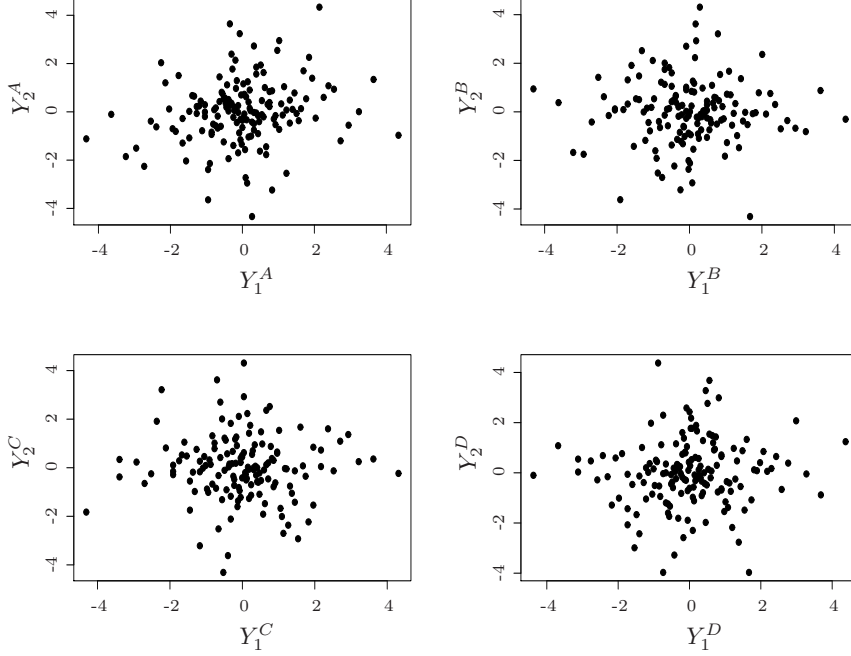


Figure 2: Scatterplots of  $Y_2^j$  against  $Y_1^j$ , for  $j = A, \dots, D$

parameter space to impose the stochastic ordering assumption between all dose levels is shown in Figure 3 via the conditional quantile estimates obtained from the SO model. For the SO model we selected  $v$  to be 5, the 99.7% quantile of the Laplace distribution. The imposition of the ordering constraints induces changes in all conditional quantile estimates which satisfy the ordering assumption above the conditioning level 5. The most important change in the quantile estimates is observed for dose  $A$  which are considerably smaller than the HT estimates, when  $q = 1$ .

### 5.3 Prediction

In this section, the main focus is placed on the prediction of joint elevations of ALT and TBL. As stated by FDA (2008) and mentioned earlier in the Introduction, DILI is associated with ALT and TBL exceeding the  $3 \times \text{ULN}_{\text{ALT}}$  and  $2 \times \text{ULN}_{\text{TBL}}$  respectively. For ALT, the ULN is taken to be 36 units/litre and for TBL is 21  $\mu\text{mol/litre}$ . Let  $p^j(x, y)$  be the joint survival probability

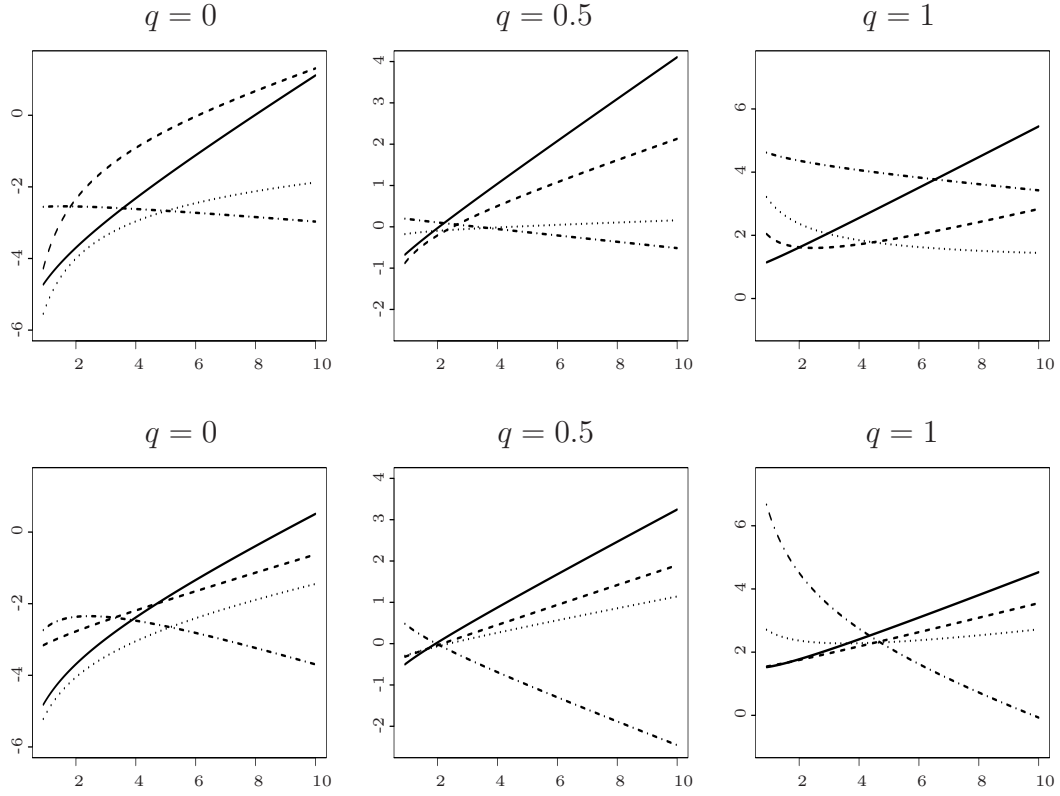


Figure 3: Conditional quantile estimates  $y_{2|1}(q)$  of  $Y_2^j|Y_1^j = x$  under HT (first row) and SO models (second row). The line types correspond to  $\cdots$  dose A,  $\cdots\cdots$  dose B,  $-\cdot-\cdot-$  dose C,  $\text{—}$  dose D. First, second and third column show the minimum ( $q = 0$ ), median ( $q = 0.5$ ) and maximum ( $q = 1$ ) conditional quantiles, respectively.

of  $\{\text{ALT} > x \cap \text{TBL} > y\}$  at dose level  $j = A, B, C$  or  $D$ , i.e.,

$$p^j(x, y) := \mathbb{P}(V_{1,P}^j > x, V_{2,P}^j > y), \quad x, y \in \mathbb{R}. \quad (12)$$

To estimate the survival probability (12) we follow the approach of Southworth and Heffernan (2012b), also mentioned earlier in Section 5.1 and simulate  $N = 200000$  post-baseline samples. For each dose level,  $N$  baseline samples are generated from the semi-parametric model (2) and are subsequently combined with  $N$  generated residual samples from the SO constrained Heffernan and Tawn (2004) model in equation (11), with  $\gamma_i^j$  and  $\delta_i^j$  replaced by their corresponding maximum likelihood estimates, to produce simulated samples for the log-post-baseline variable  $W_{i,P}^j$ . The simulated sample is then back-transformed to its original scale and the survival probability (12) is estimated empirically. To assess the uncertainty of the estimates, this procedure is repeated 1000 times and 95% equal-tail confidence intervals are obtained from the bootstrap distribution of each estimate.

Figure 4 shows the estimated survival probabilities  $p^j(x, y)$  for  $x = 3 \times \text{ULN}_{\text{ALT}}$  and variable  $y$ . For comparisons, estimates are reported from the SO and HT models. The imposition of the constraints induces changes in all estimates. In particular, the survival probability estimates from the SO model are lower than the HT model for all doses, especially in the region  $\{y : y < 30\}$ . This behaviour also implies changes in the upper tail and in the joint region of DILI, i.e., when  $x = 3 \times \text{ULN}_{\text{ALT}}$  and  $x = 2 \times \text{ULN}_{\text{TBL}}$ .

To assess the effect of the stochastic ordering constraints in the joint upper tail of ALT and TBL we study in Figure 5 the Monte Carlo estimates of extreme quantiles  $y_T^j$  defined by

$$y_T^j : \mathbb{P}(V_{1,P}^j > 3 \times \text{ULN}_{\text{ALT}}, V_{2,P}^j > y_T^j) = 1/T,$$

for large  $T$ . The level  $y_T^j$  is more commonly referred to as the return level associated with patient return period  $T$  and is expected to be exceeded, jointly with  $\{V_{1,P}^j > 3 \times \text{ULN}_{\text{ALT}}\}$ , once in  $T$  patients.

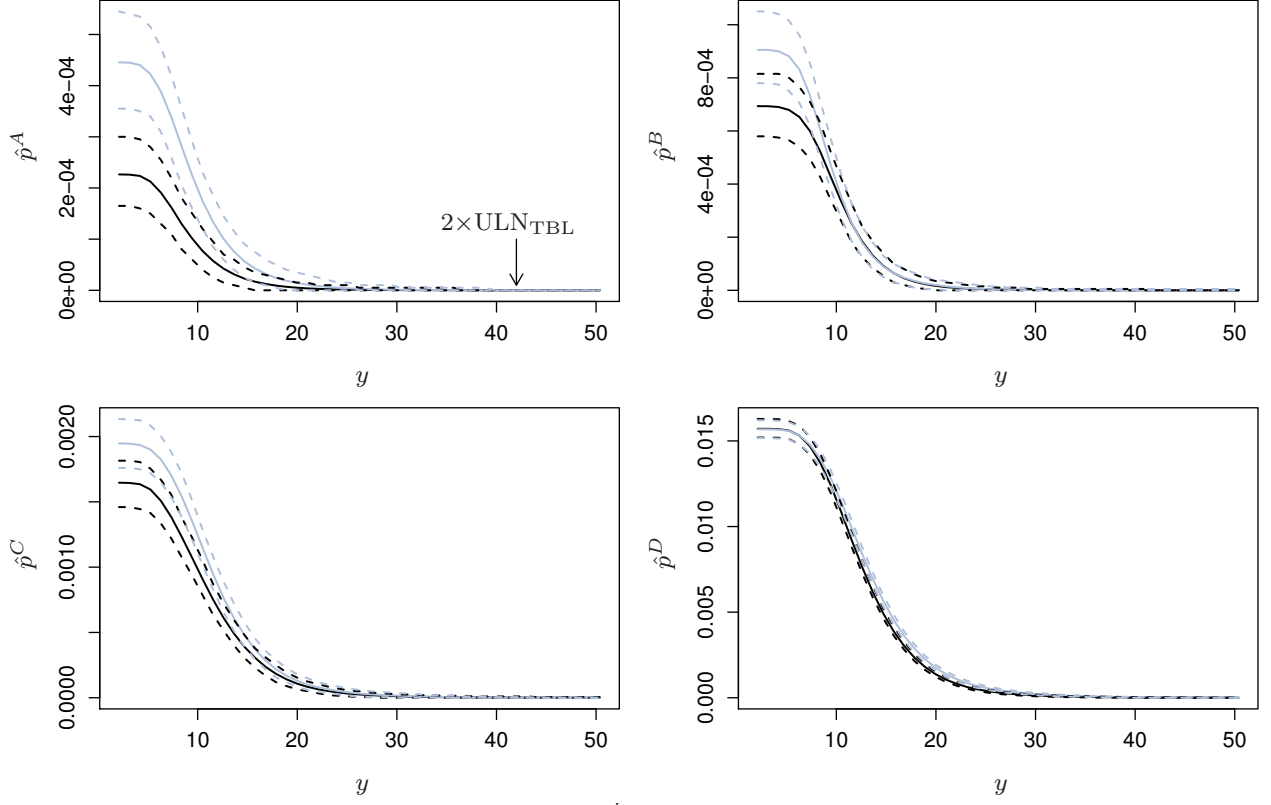


Figure 4: Estimated survival probabilities  $\hat{p}^j(x, y)$  for all doses with  $x = 3 \times \text{ULN}_{\text{ALT}}$  and variable  $y$ . Solid and dashed lines correspond the point-wise estimates and their 95% confidence intervals, respectively. Black and grey colour shows estimates from the SO model and HT models, respectively.

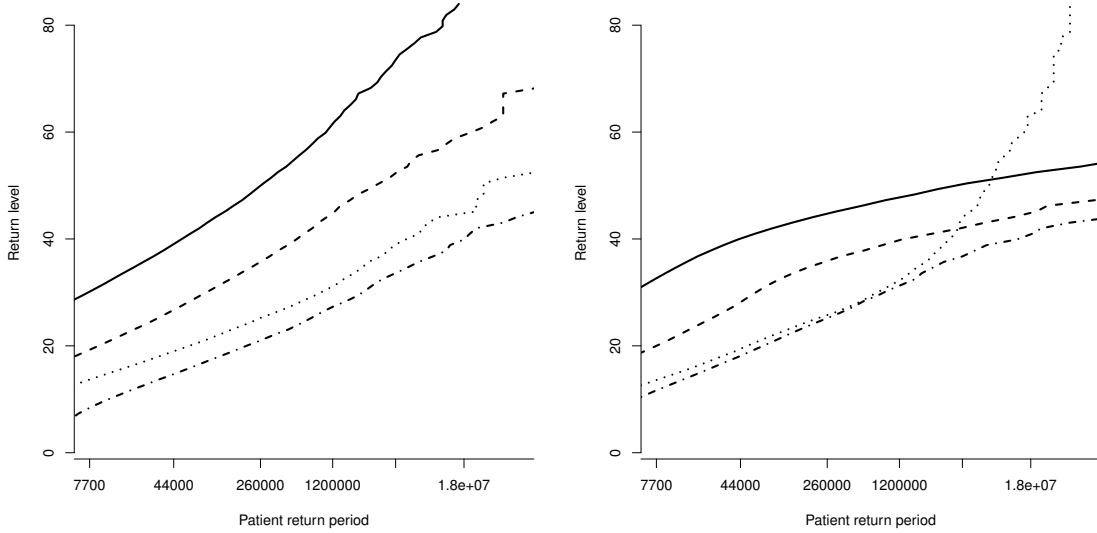


Figure 5: Monte Carlo return level estimates of return levels  $\hat{y}_T^j$  plotted against the patient return period  $T$  on the logarithmic scale. Estimates are obtained from the SO (left) and HT (right) models and the line types are as in Figure 3 and correspond to the dose levels.

The figure indicates a lack of ordering in the estimates obtained from the HT model which stems from the behaviour of the return level estimates of dose  $B$  that are unexpectedly higher than doses  $C$  and  $D$  for large patient return periods. On the other hand, the return level estimates obtained from the SO model satisfy the ordering constraints and are larger than the HT estimates. On this note, we also found smaller return level estimates than those predicted by the HT and SO models when samples of the baseline variables were generated via resampling from the observed samples and not from the semi-parametric model (2). This is attributed to the fact that the resampling approach does not allow for extrapolation in the baseline variables.

We view this lack of ordering as a by-product of sampling variability that arises in small samples and thus the SO model has corrected for this.

## Acknowledgments

I. Papastathopoulos acknowledges funding from AstraZeneca and EPSRC. We would particularly like to thank Harry Southworth of AstraZeneca for helpful discussions and constructive comments on the analysis of the pharmaceutical data.

## References

- Abbas, S., Poon, S.-H. and Tawn, J. A. (2011), ‘Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX.’, *Journal of Business and Economic Statistics*.
- Box, G. E. P. and Cox, R. D. (1964), ‘An analysis of transformations.’, *J. Roy. Statist. Soc., B* **26**, 211–252.
- Coles, S. G. and Tawn, J. A. (1994), ‘Statistical methods for multivariate extremes: an application to structural design (with discussion)’, *Appl. Statist.* **43**, 1–48.
- Davison, A. C. and Smith, R. L. (1990), ‘Models for exceedances over high thresholds’, *J. Roy. Statist. Soc., B* **52**, 393–442.
- Eastoe, E. F. and Tawn, J. A. (2012), ‘The distribution for the cluster maxima of exceedances of sub-asymptotic thresholds’, *Biometrika* **99**, 43–55.

- FDA (2008), *Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation*.
- Heffernan, J. E. and Tawn, J. A. (2004), ‘A conditional approach for multivariate extreme values (with discussion)’, *J. Roy. Statist. Soc., B* **66**, 1–34.
- Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- Keef, C., Papastathopoulos, I. and Tawn, J. A. (2012), ‘Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model’, *J. Mult. Anal.* under revision.
- Keef, C., Svenson, C. and Tawn, J. A. (2009), ‘Spatial dependence in extreme river flows and precipitation for Great Britain’, *J. Hydrology* **378**, 240–252.
- Koenker, R. and Bassett, Jr., G. (1978), ‘Regression quantiles’, *Econometrica* **46**(1), 33–50.
- Ledford, A. W. and Tawn, J. A. (1996), ‘Statistics for near independence in multivariate extreme values’, *Biometrika* **83**, 169–187.
- Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer Series in Statistics, second edn, Springer, New York.
- Papastathopoulos, I. and Tawn, J. A. (2012), ‘Extended generalised Pareto models for tail estimation’, *Statist. Plann. and Inf.* **143**, 131–143.
- Paulo, M., van der Voet, H., Wood, J., Marion, G. and van Klaveren, J. (2006), ‘Analysis of multivariate extreme intakes of food chemicals’, *Food and Chemical Toxicology* **44**(7), 994–1005.
- Southworth, H. and Heffernan, J. E. (2012a), ‘Extreme value modelling of laboratory safety data from clinical studies’, *Pharmaceut. Statist.* doi: 10.1002/pst.1510.
- Southworth, H. and Heffernan, J. E. (2012b), ‘Multivariate extreme value modelling of laboratory safety data from clinical studies’, *Pharmaceut. Statist.* doi: 10.1002/pst.1531.
- Southworth, H. and Heffernan, J. E. (2012c), *texmex: Threshold exceedences and multivariate extremes*. R package version 1.3. **URL:** <http://CRAN.R-project.org/package=texmex>

Yu, K. and Moyeed, R. A. (2001), ‘Bayesian quantile regression’, *Statist. Probab. Lett.* **54**(4), 437–447.